2009-08

Pseudospectral Optimal Control on Arbitrary Grids

Gong, Qi

AAS/AIAA Astrodynamics Specialist Conference, Pittsburgh, PA, August 9-13, 2009
http://hdl.handle.net/10945/40333

AAS/AIAA Astrodynamics Specialist Conference, Pittsburgh, PA, August 9-13, 2009

AAS 09-405

PSEUDOSPECTRAL OPTIMAL CONTROL ON ARBITRARY GRIDS

Qi Gong,* I. Michael Ross† and Fariba Fahroo ‡

In advancing our prior work on a unified theory for pseudospectral (PS) optimal control, we present new results for PS methods over arbitrary grids. These results provide a way to compare performances among different PS methods and suggest guidelines to choose the proper grids and discretization approaches for solving optimal control problems. The new unified ideas reveal hidden properties of different types of PS methods and pave the way to construct more efficient algorithms for solving different types of optimal control problems. Our computational framework is not based on any particular choice of orthogonal polynomials; therefore, the approach unifies Legendre and Chebyshev PS methods under one framework.

INTRODUCTION

In 2007, a front page article in *SIAM News*[1] announced that pseudospectral (PS) optimal control was successfully used to maneuver the International Space Station. This breakthrough was preceded by a large body of work on Legendre PS methods[2–10] Since then, a number of ground and flight experiments have been conducted on various projects around the world in both academia and industry. Such successes have lead practitioners to demand more performance and higher versatility of PS methods. Motivated by such requirements and a fundamental quest to unify various PS methods, new ideas were put forth by Fahroo and Ross in Refs. [11–13]. In this paper, we further these ideas to reveal a number of fundamental properties of PS methods that were previously hidden due to the special nature of previous methods.

Pseudospectral methods are a family of computational methods that can be characterized according to the underlying orthogonal polynomials, e.g., Legendre polynomials, Chebyshev polynomials, etc.; and types of quadrature nodes, e.g., Gauss-Lobatto, Gauss-Radau or pure Gauss.[11] To date, only the Legendre PS method has been mathematically proved to guarantee the feasibility, consistency and convergence of the approximations.[6,14–16] Efforts to improve the Legendre PS methods by using either other polynomials or point distributions have not yet resulted in any rigorous framework for convergence of these approximations.[11,17]

The purpose of this paper is to build a foundation for PS methods on arbitrary grids. There are several important reasons to consider PS methods on arbitrary grids. First, it provides a unified framework for the analysis of different PS methods. Most PS methods use quadrature type of nodes determined by the roots or the extrema of orthogonal polynomials. Such quadrature nodes share a common property of dense distribution around the end points, which is a signature property of PS methods; however, the exact distribution of the nodes depends on the related orthogonal polynomials. Since different orthogonal polynomials have different properties with respect to interpolation, integration, and discretization of the optimal control problem, the approximations of

*Assistant professor, Dept. of Applied Mathematics and Statistics, Univ. of California, Santa Cruz, CA, 95064, qigong@soe.ucsc.edu
†Professor, Dept. of Mech. and Astro. Engr., Naval Postgraduate School, Monterey, CA, 93943, imross@nps.edu
‡Professor, Dept. of Applied Mathematics, Naval Postgraduate School, Monterey, CA, 93943, ffahroo@nps.edu

the states, control and costates based on these various methods could have different outcomes. A specific challenge that is addressed in this paper is that for polynomials orthogonal with respect to a non-constant function such as Chebyshev polynomials, the development of a covector mapping theorem is nontrivial.

In this paper we present new results for PS methods over arbitrary grids. These results provide a way to compare performances among different PS methods and suggest guidelines for choosing the proper grids and discretization approaches for solving optimal control problems. For instance, by using weighted interpolants, Fahroo and Ross[11] developed a unified framework for Legendre PS methods for all Gaussian grid points: Gauss-Lobatto, Gauss-Radau and Gauss. This unified point of view provided a clear explanation why the choice of quadrature nodes should depend on the boundary conditions of the optimal control problems. This paper generalizes such results on arbitrary grids and reveals hidden properties of different types of PS methods and paves the way to construct more efficient algorithms for solving different types of optimal control problems. Our analysis is not based on any particular choice of orthogonal polynomials; therefore, the approach unifies Legendre and Chebyshev PS methods under one umbrella. A new set of primal-only closure condition is developed for the purpose of costate computation. This approach is applicable to different PS methods including both Legendre and Chebyshev PS methods. Analysis on feasibility and consistency of the PS methods on arbitrary grids is also provided. Various examples are presented in the paper to demonstrate the performance of the proposed algorithms.

FOUNDATIONS OF PSEUDOSPECTRAL OPTIMAL CONTROL

In this section, we lay down the fundamental ideas of pseudospectral optimal control. To this end, consider the following nonlinear constrained optimal control problem.

$$(B)\begin{cases} \text{Minimize} & J[x(\cdot),u(\cdot)] = E(x(-1),x(1)) + \displaystyle\int_{-1}^{1} F(x(t),u(t))\,dt \\ \text{Subject to} & \dot{x}(t) = f(x(t),u(t)) \\ & e(x(-1),x(1)) = 0 \\ & h(x(t),u(t)) \leq 0 \end{cases}$$

That is, we have to find the state-control function pair, $t \mapsto (x,u) \in \mathbb{R}^{N_x} \times \mathbb{R}^{N_u}$, that solves Problem B. It is assumed that $F : \mathbb{R}^{N_x} \times \mathbb{R}^{N_u} \to \mathbb{R}$, $E : \mathbb{R}^{N_x} \times \mathbb{R}^{N_x} \to \mathbb{R}$, $f : \mathbb{R}^{N_x} \times \mathbb{R}^{N_u} \to \mathbb{R}^{N_x}$, $e : \mathbb{R}^{N_x} \times \mathbb{R}^{N_x} \to \mathbb{R}^{N_e}$, and $h : \mathbb{R}^{N_x} \times \mathbb{R}^{N_u} \to \mathbb{R}^{N_h}$, are continuous with respect to their arguments and their gradients are Lipschitz continuous over the domain. Note that, the continuity of the vector fields does not exclude the discontinuous optimal control. It is well known that a smooth optimal control problem may yield discontinuous solutions like bang-bang control. In order to apply the first order necessary conditions, appropriate constraint qualifications are implicitly assumed so that the first order necessary conditions hold. It should be noted that using a simple time domain transformation, the results would hold for problems on $t \in [a,b]$ and time free problems can be easily handled as well.

The first fundamental step behind a PS method is to design Problem B^N which is expected to be an approximation to Problem B in such a manner that as $N \to \infty$, the solution to Problem B^N converges to the solution of Problem B. We note once again that an improper selection of grid points can lead to divergence; hence, the design of Problem B^N is a crucial first step to formulating a proper PS method.

Differentiation and Integration over Arbitrary Grids

Because, Problem B has two end points fixed at -1 and 1, we choose the initial and the final grid points to be $t_0 = -1$ and $t_N = 1$. Between two end points, the grids are arbitrary and denoted as $-1 < t_1 < t_2 < \cdots < t_{N-1} < 1$. Let $y(t)$ be an arbitrary scalar function. In any PS method, the basic idea is to approximate $y(t)$ by an interpolating basis functions. We choose polynomial basis functions because of Weierstrass' approximation theorem which states (roughly) that any continuous function can be approximated by a polynomial of sufficiently high order. For any given grid, we define,

$$y^N(t) = \sum_{j=0}^{N} \phi_j(t) y_j, \ -1 \le t \le 1 \tag{1}$$

where $\phi_j(t)$ is the $Nth-$ order Lagrange interpolating polynomial that satisfies the relationship $\phi_j(t_k) = \delta_{jk}$. This implies that

$$y_j = y^N(t_j), \ \ j = 0, ...N. \tag{2}$$

Note that this *does not imply* that $y(t) \approx y^N(t)$; in fact, $y(t)$ can be very far from $y^N(t)$ even when $y(t)$ is known! This famous statement, due to Runge, known as the Runge phenomenon is illustrated in Fig. 1.

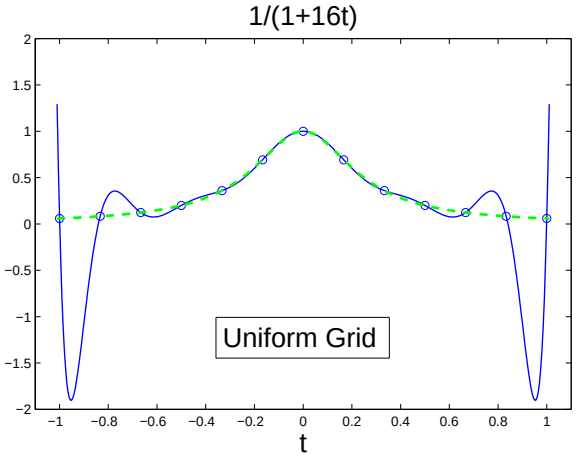

Figure 1 Runge's famous counter example disproving presumed convergence of interpolating polynomials over a uniform grid.

An expression for the Lagrange polynomial can be written as[20]

$$\phi_j(t) = \frac{g_N(t)}{g'_N(t_j)(t - t_j)}, \ g_N(t) = \prod_{j=0}^{N} (t - t_j). \tag{3}$$

One important tenant of PS approximation of functions is that differentiation of the approximated functions can be performed by differentiation of the interpolating polynomial,

$$\frac{dy^N(t)}{dt} = \sum_{j=0}^{N} y_j \phi'_j(t)$$

For computation, only the values of the derivative at the nodes t_i are required, hence we have,

$$\left.\frac{dy^N(t)}{dt}\right|_{t_i} = \sum_{j=0}^{N} D_{ij} y_j \tag{4}$$

where D_{ij} is the differentiation matrix defined by,

$$D_{ij} := \left.\frac{d\phi_j(t)}{dt}\right|_{t=t_i} \tag{5}$$

From Eq. (3), the differentiation matrix, $D_{ij} = \phi_j'(t_i)$, can be written in the following explicit form,

$$D_{ij} = \begin{cases} \dfrac{g_N'(t_i)}{g_N'(t_j)} \dfrac{1}{(t_i - t_j)}, & i \neq j \\ \\ \dfrac{g_N''(t_i)}{2g_N'(t_i)}, & i = j \end{cases} \tag{6}$$

The above equations are the general representations of the derivative of the Lagrange polynomials evaluated at arbitrary interpolation nodes. From Runge's counter example, it is clear that an improper selection of the grid points can lead to disastrous consequences. In fact, a uniform distribution of grid points is the worst possible choice for polynomial interpolation and hence differentiation.

Similar to differentiation, integration can also be approximated using Lagrange interpolating polynomial. Let,

$$I(y^N) := \int_{-1}^{+1} y^N(t)\, dt$$

Substituting (1) above, we get,

$$I(y^N) := \sum_{j=0}^{N} y_j w_j \tag{7}$$

where w_j are the weights given by

$$w_j \quad = \quad \int_{-1}^{1} \phi_k(t) dt \qquad j = 0, 1, \cdots, N \tag{8}$$

Remark 1 *Once the choice of grids is fixed, the differentiation matrix and integration weights are completely determined.*

Remark 2 *With $N + 1$ nodes, this integration scheme is exact for any polynomial of order N. Indeed, let $p^N(t)$ be any polynomial of order N or less,*

$$\begin{aligned} \int_{-1}^{1} p^N(t) dt &= \int_{-1}^{1} \sum_{k=0}^{N} p^N(t_k) \phi_k(t) dt \\ &= \sum_{k=0}^{N} p^N(t_k) \int_{-1}^{1} \phi_k(t) dt \\ &= \sum_{k=0}^{N} p^N(t_k) w_k \end{aligned}$$

Remark 3 *In the case of Legendre-Gauss-Lobatto nodes, the differentiation matrix and the weights are given by*

$$D_{jk} = \begin{cases} \frac{L_N(t_j)}{L_N(t_k)}\frac{1}{t_j - t_k}, & \text{if } j \neq k; \\ -\frac{N(N+1)}{4}, & \text{if } j = k = 0; \\ \frac{N(N+1)}{4}, & \text{if } j = k = N; \\ 0, & \text{otherwise} \end{cases}$$

$$w_k = \frac{2}{N(N+1)}\frac{1}{[L_N(t_k)]^2}, \qquad k = 0, 1, \cdots, N$$

where $L_N(t)$ is the N-th order Legendre polynomial. If Chebyshev-Gauss-Lobatto points are chosen, the differentiation matrix is

$$D_{jk} = \begin{cases} (c_j/c_k)\left[(-1)^{j+k}/(t_j - t_k)\right], & \text{if } j \neq k; \\ t_k/(2 - 2t_k^2), & \text{if } 1 \leq j = k \leq N-1; \\ -(2N^2+1)/6, & \text{if } j = k = 0; \\ (2N^2+1)/6, & \text{if } j = k = N \end{cases}$$

and the weights are Clenshaw-Curtis weights. For N even, the weights are $w_0 = w_N = 1/(N^2-1)$, and

$$w_s = w_{N-s} = \frac{4}{N}\sum_{j=0}^{N/2''}\frac{1}{1-4j^2}\cos\left(\frac{2\pi js}{N}\right), \; s = 1, 2, \ldots, N/2$$

For N odd, the weights are given by $w_0 = w_N = 1/N^2$, and

$$w_s = w_{N-s} = \frac{4}{N}\sum_{j=0}^{(N-1)/2''}\frac{1}{1-4j^2}\cos\left(\frac{2\pi js}{N}\right),$$
$$s = 1, \ldots, (N-1)/2$$

The double prime in the preceding formulas means that the first and the last elements have to be halved.

Construction of Problem B^N

From the fundamentals of the preceding subsection, it is clear that once the grid is chosen, differentiation and integration are naturally defined. This leads to an automatic definition of Problem B^N as follows: Define $x_i^N(t)$ as

$$x_i(t) \approx \; x_i^N(t) = \sum_{k=0}^{N}\bar{x}_i^k\phi_k(t),$$

where $x_i(t)$ is the i-th component of the state $x(t)$. Let

$$\bar{x}^k = \left[\; \bar{x}_1^k, \quad \bar{x}_2^k, \quad \cdots, \quad \bar{x}_{N_x}^k \;\right]^T$$

and

$$\bar{x}_i = \left[\begin{array}{cccc} \bar{x}_i^0 & \bar{x}_i^1 & \cdots & \bar{x}_i^N \end{array} \right]$$

In this notation, the discrete variables are denoted by letters with an upper bar, such as $\bar{x}_i^k$ and $\bar{u}_i^k$. If k in the superscript and/or i in the subscript are missing, it represents the corresponding vector or matrix in which the indices run from minimum to maximum. For example, let

$$\bar{x} \quad = \quad \left[\begin{array}{cccc} \bar{x}_1^0 & \bar{x}_1^1 & \cdots & \bar{x}_1^N \\ \bar{x}_2^0 & \bar{x}_2^1 & \cdots & \bar{x}_2^N \\ \vdots & \vdots & \vdots & \vdots \\ \bar{x}_{N_x}^0 & \bar{x}_{N_x}^1 & \cdots & \bar{x}_{N_x}^N \end{array} \right]$$

$\bar{x}_i$ is the ith row of $\bar{x}$, which represents the purported discrete approximation of the ith component, $x_i(t)$, at all nodes; and $\bar{x}^i$ is the ith column of $\bar{x}$, which represents the approximation of the state, $x(t)$, at ith node. Thus the derivative of $x_i^N(t)$ at the node t_k can be computed by the following matrix multiplication

$$\left[\begin{array}{cccc} \dot{x}_i^N(t_0), & \dot{x}_i^N(t_1), & \cdots & \dot{x}_i^N(t_N) \end{array} \right] = \bar{x}_i \cdot D^T$$

where the $(N+1) \times (N+1)$ differentiation matrix D is defined in previous section. This implies that the differential equation can be enforced over all grids points by collocation, i.e.,

$$\sum_{j=0}^{N} D_{kj} \bar{x}^j \quad = \quad f(\bar{x}^k, \bar{u}^k), \qquad k = 0, 1, \cdots, N$$

Similarly the cost functional is discretized as

$$J[x(\cdot), u(\cdot)] \approx \bar{J}^N(\bar{x}^N, \bar{u}^N) = \sum_{k=0}^{N} F(\bar{x}^k, \bar{u}^k) w_k + E(\bar{x}^0, \bar{x}^N)$$

This generates the following construction for Problem B^N:

Problem $\mathbf{B^N}$: Let $\mathbb{X}$ and $\mathbb{U}$ be two compact sets representing the search region. Find $\bar{x} \in \mathbb{X}$ and $\bar{u} \in \mathbb{U}$ that minimize

$$\bar{J}^N(\bar{x}, \bar{u}) \quad = \quad \sum_{k=0}^{N} F(\bar{x}^k, \bar{u}^k) w_k + E(\bar{x}^0, \bar{x}^N)$$

subject to the discrete dynamics

$$-\sum_{j=0}^{N} D_{kj} \bar{x}^j + f(\bar{x}^k, \bar{u}^k) \quad = \quad 0$$

end point constraints $e(\bar{x}^0, \bar{x}^N) = 0$ and path constraints $h(\bar{x}^k, \bar{u}^k) \leq 0$ for all $k = 0, 1, \ldots, N$.

EXISTENCE AND CONSISTENCY

In this section, we will establish theoretical foundation on the feasibility and the consistency of PS methods. The analysis shows that the presented PS discretization always results in a feasible nonlinear optimization problem. Furthermore, we identify some conditions under which the limit point of the discretized optimal solution is guaranteed to be the optimal solution of the continuous Problem B. These conditions provide some guideline for the choice of the grids. The proofs of these results follow our prior works on Legendre PS method[6,14] and hence omitted.

Feasibility of Problem B^N

When the constraints are discretized into exact equations as in Problem B^N, it is not surprising that the constraint set may contain no solution. This is true even in the case of Euler discretization. The following is a counter example from Ref. [6].

Example 1 Consider the linear system

$$\begin{aligned} \dot{x}_1 &= x_1 + u \\ \dot{x}_2 &= x_2 + u \end{aligned} \tag{9}$$

Its PS discretization is

$$\begin{aligned} D \begin{pmatrix} \bar{x}_1^0 \\ \vdots \\ \bar{x}_1^N \end{pmatrix} &= \begin{pmatrix} \bar{x}_1^0 \\ \vdots \\ \bar{x}_1^N \end{pmatrix} + \begin{pmatrix} \bar{u}^0 \\ \vdots \\ \bar{u}^N \end{pmatrix} \\ D \begin{pmatrix} \bar{x}_2^0 \\ \vdots \\ \bar{x}_2^N \end{pmatrix} &= \begin{pmatrix} \bar{x}_2^0 \\ \vdots \\ \bar{x}_2^N \end{pmatrix} + \begin{pmatrix} \bar{u}^0 \\ \vdots \\ \bar{u}^N \end{pmatrix} \end{aligned}$$

Therefore

$$(D-I) \begin{pmatrix} \bar{x}_1^0 \\ \vdots \\ \bar{x}_1^N \end{pmatrix} = (D-I) \begin{pmatrix} \bar{x}_2^0 \\ \vdots \\ \bar{x}_2^N \end{pmatrix}$$

It is easy to show that for any grids, D is always nilpotent. Therefore, $(D-I)$ is nonsingular. Hence, $(\bar{x}_1^0, \ldots, \bar{x}_1^N) = (\bar{x}_2^0, \ldots, \bar{x}_2^N)$. It implies that, if the initial condition is such that $\bar{x}_1^0 \neq \bar{x}_2^0$, the discretized dynamics with arbitrary initial conditions has no solution, although a continuous solution satisfying (9) always exists for any given initial condition.

Similar to Euler methods, the remedy lies in proper relaxation principles.[18,19] This implies that the constraints in Problem B^N are relaxed to

$$\begin{aligned} \left|\left| -\sum_{j=0}^{N} D_{kj} \bar{x}^j + f(\bar{x}^k, \bar{u}^k) \right|\right|_\infty &\leq \delta^N \\ \left|\left| e(\bar{x}^0, \bar{x}^N) \right|\right|_\infty &\leq \delta^N \\ h(\bar{x}^k, \bar{u}^k) &\leq \delta^N \cdot \mathbf{1} \end{aligned}$$

where $\mathbf{1} = [1, \cdots, 1]^T$ and δ^N is a small positive number that depends upon N. This mathematics has a very practical counterpart: when Problem B^N is solved computationally, it is impossible for the constraints to be satisfied exactly. Because the constraints are always relaxed by some feasibility tolerance, δ^N is, in fact the feasibility tolerance. When the continuous solution is sufficiently smooth, it can be proved that the relaxed constraints are always feasible. To understand this statement, we need the following definition:

Definition 1 *A function, $\xi : [-1, 1] \to \mathbb{R}$ belongs to Sobolev spaces,[20] $W^{m,p}$, if its j-th weak derivative, $\xi^{(j)}$, lies in $L^p[-1, 1]$ for all $0 \leq j \leq m$.*

Theorem 1 *Given any feasible solution, $t \mapsto (x, u)$, for Problem B, suppose $x(\cdot) \in W^{m,\infty}$ with $m \geq 2$. Then, there exists a positive integer N_1 such that, for all $N > N_1$, Problem $\mathrm{B^N}$ has a feasible solution with the feasibility tolerance be $\delta^N = (N-1)^{3/2-m}$. Furthermore, the feasible solution satisfies $\bar{u}^k = u(t_k)$ and*

$$\|\bar{x}^k - x(t_k)\|_\infty \quad \leq \quad L(N-1)^{1-m},$$

for all $k = 0, \ldots, N$, where t_k are arbitrarily chosen grids with $t_0 = -1$, $t_N = 1$ and L is a positive constant independent of N.

Theorem 1 guarantees that the discretized optimization problem is well posed, as long as the continuous optimal control problem admits a sufficiently smooth solution.

Remark 4 *The assumption, $m \geq 2$, guarantees that*

$$\lim_{N\to\infty} \delta^N = \lim_{N\to\infty} (N-1)^{3/2-m} = 0.$$

This condition is needed to establish the convergence results. Also, because of this property, for any fixed constant δ, the discretized Problem B^N is always feasible with sufficiently large number of nodes.

Remark 5 *If $x(\cdot) \in C^1$ and $\dot{x}(\cdot)$ has a bounded derivative everywhere except for finitely many points, then it is clear that $x(t) \in W^{2,\infty}$. From Sobolev's Imbedding Theorems,[20] any function $x(\cdot) \in W^{m,\infty}$, $m \geq 2$ must have continuous $(m-1)$-th order classical derivatives (on $[-1, 1]$). Therefore, the condition, $x(\cdot) \in W^{m,\infty}$, $m \geq 2$, requires that the state trajectory, $x(\cdot)$, be at least continuously differentiable, which in turn requires the control, $u(\cdot)$, be continuous. It is possible to relax the smoothness assumption by one order to include the discontinuous control.[15]*

Consistency

Let $(\bar{x}^*, \bar{u}^*)_N$, be an optimal solution to Problem $\mathrm{B^N}$; $x^N(t) \in \mathbb{R}^{N_x}$ be the N-th order interpolating polynomial of $\bar{x}^*$ and $u^N(t) \in \mathbb{R}^{N_u}$ be any interpolant of $\bar{u}^*$, i.e.

$$x^N(t) = \sum_{k=0}^{N} \bar{x}^{k*} \phi_k(t), \qquad u^N(t) = \sum_{k=0}^{N} \bar{u}^{k*} \psi_k(t)$$

where $\phi_k(t)$ is the Lagrange interpolating polynomial defined by (3) and $\psi_k(t)$ is any continuous function such that $\psi_k(t_j) = 1$, if $k = j$ and $\psi_k(t_j) = 0$, if $k \neq j$. Note that $u^N(t)$ is not necessarily a polynomial, but an interpolating function.

Now consider a sequence of Problems $\mathrm{B^N}$ with N increasing from N_1 to infinity. The nodes distribution as N turns to infinity is assumed to be dense. Correspondingly, we get a sequence of discrete optimal solutions $\{\bar{x}^*, \bar{u}^*\}_{N=N_1}^{\infty}$ and their interpolating function sequence $\{x^N(t), u^N(t)\}_{N=N_1}^{\infty}$. The following result shows that, under some assumptions, the limit points of this function sequence will be the optimal solution to the original continuous optimal control problem.

Assumption 1 *Let the feasibility tolerance $\delta = (N-1)^{3/2-m}$ as in Theorem 1; $\{\bar{x}^*, \bar{u}^*\}_{N=N_1}^{\infty}$ be a sequence of optimal solutions to Problem $\mathrm{B^N}$ and $\{x^N(t), u^N(t)\}_{N=N_1}^{\infty}$ be their interpolating function sequence. Suppose $\{\bar{x}^{0*}, \dot{x}^N(\cdot), u^N(\cdot)\}_{N=N_1}^{\infty}$ has a subsequence that uniformly converges to $(x_0^{\infty}, q(\cdot), u^{\infty}(\cdot))$, where $q(t)$ and $u^{\infty}(t)$ are continuous on $t \in [-1, 1]$.*

Assumption 2 *For all chosen grids sequence, the weights defined in (8) are positive; and for any Riemann integrable function $\xi(t)$,*

$$\int_{-1}^{1} \xi(t)dt \quad = \quad \lim_{N\to\infty} \sum_{k=0}^{N} \xi(t_k) w_k$$

Theorem 2 *Under Assumptions 1 and 2, the limit function $u^{\infty}(t)$ is an optimal control to the original continuous Problem B, and $x^{\infty}(t) = \int_{-1}^{t} q(\tau)d\tau + x_0^{\infty}$ is the corresponding optimal trajectory.*

This result demonstrates that Problem B^N is indeed a consistent approximation[21] to the continuous optimal control Problem B. *Under Assumption 2, if the optimal solution of the discrete-time Problem B^N converges as N increases, then the limit point must be an optimal solution of the continuous-time Problem B*. Under some extra conditions, it is also possible to guarantee the existence of a convergent subsequence.[16]

Remark 6 *Assumption 2 is critical to guarantee the consistency. This assumption imposes requirements on the choice of the grids, since the weights are completely determined by the nodes. It is well- known that the uniform distribution of grids indeed does not satisfy this assumption; therefore should not be used for the proposed scheme. On the contrary, quadrature types of nodes like Legendre-Gauss-Lobatto or Chebyshev-Gauss-Lobatto nodes do satisfy this condition. This assumption also indicates that the choice of grids should make the approximation of the integration as accurate as possible. Later we will show that the accuracy of the numerical integration also affects the computation of the costate.*

DUAL SPACE AND COSTATE COMPUTATION

Dualization

In this section, we examine the necessary conditions of the presented pseudospectral method and propose a new primal-only closure condition for the costate computation. From Pontryagin's Principle, Problem B generates a boundary value problem that can be summarized as Problem B^{λ}:

Problem B^{λ}: If $(x(t), u(t))$ is the optimal solution to Problem B, then there exist $(\lambda(t), \mu(t), \nu)$ such that

$$\begin{aligned}
\dot{x} &= f(x,u) \\
\dot{\lambda} &= -F_x(x,u) - f_x^T(x,u)\lambda - h_x^T(x,u)\mu(t) \\
0 &= F_u(x,u) + f_u^T(x,u)\lambda + h_u^T(x,u)\mu(t) \\
0 &= e(x(1), x(-1)) \\
0 &\geq h(x,u) \\
0 &= \mu(t)h(x(t),u(t)), \qquad \mu(t) \geq 0 \\
\lambda(-1) &= -E_{x(-1)}(x(-1),x(1)) - e_{x(-1)}^T(x(-1),x(1))\nu \\
\lambda(1) &= E_{x(1)}(x(-1),x(1)) + e_{x(1)}^T(x(-1),x(1))\nu
\end{aligned}$$

The presented PS discretization of this problem generates Problem $B^{\lambda N}$:

Problem $B^{\lambda N}$: Find $(\bar{x}, \bar{u}, \bar{\lambda}, \bar{\mu}, \bar{\nu})$, such that

$$\begin{aligned}
&-\sum_{j=0}^{N} D_{kj}\bar{x}^j + f(\bar{x}^k, \bar{u}^k) = 0 \\
&\sum_{j=0}^{N} D_{kj}\bar{\lambda}^j = -f_x^T(\bar{x}^k,\bar{u}^k)\bar{\lambda}^k - F_x(\bar{x}^k,\bar{u}^k)w_k - h_x^T(\bar{x}^k,\bar{u}^k)\bar{\mu}^k \\
&F_u(\bar{x}^k,\bar{u}^k)w_k + f_u^T(\bar{x}^k,\bar{u}^k)\bar{\lambda}^k + h_u^T(\bar{x}^k,\bar{u}^k)\bar{\mu}^k = 0 \\
&e(\bar{x}^0, \bar{x}^N) = 0 \\
&h(\bar{x}^k,\bar{u}^k) \leq 0 \\
&\bar{\mu}^k \cdot h(\bar{x}^k,\bar{u}^k) = 0, \qquad \bar{\mu}^k \geq 0 \\
&\bar{\lambda}^0 = -\frac{\partial E}{\partial x^0}(\bar{x}^0,\bar{x}^N) - (\frac{\partial e}{\partial x^0}(\bar{x}^0,\bar{x}^N))^T\bar{\nu} \\
&\bar{\lambda}^N = \frac{\partial E}{\partial x^N}(\bar{x}^0,\bar{x}^N) + (\frac{\partial e}{\partial x^N}(\bar{x}^0,\bar{x}^N))^T\bar{\nu}
\end{aligned}$$

Consider now the necessary condition of the discretized optimization Problem B^N. To this end, we construct the Lagrangian for Problem B^N as

$$\begin{aligned}
L^N = \bar{J}^N &+ \sum_{k=0}^{N}(\bar{\lambda}^k)^T\left(-\sum_{j=0}^{N} D_{kj}\bar{x}^j + f(\bar{x}^k,\bar{u}^k)\right) \\
&+\bar{\nu}^T e(\bar{x}^0,\bar{x}^N) + \sum_{k=0}^{N}(\bar{\mu}^k)^T h(\bar{x}^k,\bar{u}^k)
\end{aligned}$$

where $\bar{\lambda}^k \in \mathbb{R}^{N_x}$, $\bar{\nu} \in \mathbb{R}^{N_e}$ and $\bar{\mu}^k \in \mathbb{R}^{N_h}$ are the Karush-Kuhn-Tucker (KKT) multipliers. The KKT conditions are summarized in the following.

Problem $B^{N\lambda}$: Find $(\bar{x}, \bar{u}, \bar{\lambda}, \bar{\mu}, \bar{\nu})$, such that

$$
\begin{aligned}
&-\sum_{j=0}^{N} D_{kj}\bar{x}^j + f(\bar{x}^k, \bar{u}^k) \;=\; 0 \\
&e(\bar{x}^0, \bar{x}^N) \;=\; 0 \\
&h(\bar{x}^k, \bar{u}^k) \;\leq\; 0 \\
&-\sum_{j=0}^{N} D_{jk}\bar{\lambda}^j + f_x^T(\bar{x}^k, \bar{u}^k)\bar{\lambda}^k + F_x(\bar{x}^k, \bar{u}^k)w_k + h_x^T(\bar{x}^k, \bar{u}^k)\bar{\mu}^k + c_k \;=\; 0 \qquad (10) \\
&F_u(\bar{x}^k, \bar{u}^k)w_k + f_u^T(\bar{x}^k, \bar{u}^k)\bar{\lambda}^k + h_u^T(\bar{x}^k, \bar{u}^k)\bar{\mu}^k \;=\; 0 \\
&\bar{\mu}^k \cdot h(\bar{x}^k, \bar{u}^k) \;=\; 0, \qquad \bar{\mu}^k \;\geq\; 0
\end{aligned}
$$

where $c_i = 0$ for $1 \leq i \leq N-1$ and

$$
\begin{aligned}
c_0 &= \frac{\partial E}{\partial x^0}(\bar{x}^0, \bar{x}^N) + (\frac{\partial e}{\partial x^0}(\bar{x}^0, \bar{x}^N))^T\bar{\nu} \\
c_N &= \frac{\partial E}{\partial x^N}(\bar{x}^0, \bar{x}^N) + (\frac{\partial e}{\partial x^N}(\bar{x}^0, \bar{x}^N))^T\bar{\nu}
\end{aligned}
$$

By a simple inspection of Problem $B^{\lambda N}$, it is clear that we need

$$c_0 = 0 = c_N$$

These equations are the primal-dual closure conditions explained in detail in Refs.[7, 14]. We now show that this primal-dual closure conditions can be "dualized" to generate primal-only closure conditions.

Primal-Only Closure Conditions

The primal-dual closure conditions can be transformed to primal-only closure conditions. To generate this transformation, we modify Problem B^N as follows. Introduce new primal decision variables $\hat{x}^0 \in \mathbb{R}^{N_x}$ and $\hat{x}^N \in \mathbb{R}^{N_x}$ whose meanings will be apparent shortly. Essentially, $\hat{x}^0$ and $\hat{x}^N$ represent almost the same quantities as $\bar{x}^0$ and $\bar{x}^N$. In the discretization of the event cost, $E(x(-1), x(1))$, and event conditions, $e(-1, 1) = 0$, we use $\hat{x}^0$ and $\hat{x}^N$ instead of $\bar{x}^0$ and $\bar{x}^N$. Therefore, the discrete cost and the event conditions are

$$
\begin{aligned}
\bar{J}^N &= \sum_{k=0}^{N} F(\bar{x}^k, \bar{u}^k)w_k + E(\hat{x}^0, \hat{x}^N) \\
0 &= e(\hat{x}^0, \hat{x}^N)
\end{aligned}
$$

The discretization of the dynamical equations at any interior nodes t_k, $1 \leq k \leq N-1$, remains the same. But at two ends t_0 and t_N, the discrete dynamics is relaxed by $(\bar{x}^0 - \hat{x}^0)/w_0$ and $(\hat{x}^N - \bar{x}^N)/w_N$, i.e.,

$$
\begin{aligned}
-\sum_{j=0}^{N} D_{0j}\bar{x}^j + f(\bar{x}^0, \bar{u}^0) &= (\bar{x}^0 - \hat{x}^0)/w_0 \\
-\sum_{j=0}^{N} D_{Nj}\bar{x}^j + f(\bar{x}^N, \bar{u}^N) &= (\hat{x}^N - \bar{x}^N)/w_N
\end{aligned}
$$

Since both $(\hat{x}^0, \hat{x}^N)$ and $(\bar{x}^0, \bar{x}^N)$ represent the same quantities $(x(-1), x(1))$, we expect the relaxation factors $(\bar{x}^0 - \hat{x}^0)/w_0$ and $(\hat{x}^N - \bar{x}^N)/w_N$ to be close to zero. Under such primal-only modifications, the discretized optimal control problem is given by:

Modified Problem B^N: Find $(\bar{x}, \bar{u}, \hat{x}^0, \hat{x}^N)$ that minimize

$$J^N \;=\; \sum_{k=0}^{N} F(\bar{x}^k, \bar{u}^k) w_k + E(\hat{x}^0, \hat{x}^N)$$

subject to:

$$\begin{aligned}
&-\sum_{j=0}^{N} D_{0j}\bar{x}^j + f(\bar{x}^0, \bar{u}^0) \;=\; (\bar{x}^0 - \hat{x}^0)/w_0 \\
&-\sum_{j=0}^{N} D_{kj}\bar{x}^j + f(\bar{x}^k, \bar{u}^k) \;=\; 0, \;\; k = 1, \ldots, N-1 \\
&-\sum_{j=0}^{N} D_{Nj}\bar{x}^j + f(\bar{x}^k, \bar{u}^k) \;=\; (\hat{x}^N - \bar{x}^N)/w_N \\
&e(\hat{x}^0, \hat{x}^N) \;=\; 0 \\
&h(\bar{x}^k, \bar{u}^k) \;\leq\; 0
\end{aligned}$$

In this modified optimization problem, the extra variables $(\hat{x}^0, \hat{x}^N)$ provide two more KKT conditions

$$\begin{aligned}
\frac{\partial L}{\partial \hat{x}^0} &= \hat{\lambda}^0 + \frac{\partial E}{\partial x^0}(\hat{x}^0, \hat{x}^N) + (\frac{\partial e}{\partial x^0}(\hat{x}^0, \hat{x}^N))^T \hat{\nu} \;=\; 0 \\
\frac{\partial L}{\partial \hat{x}^N} &= -\hat{\lambda}^N + \frac{\partial E}{\partial x^N}(\hat{x}^0, \hat{x}^N) + (\frac{\partial e}{\partial x^N}(\hat{x}^0, \hat{x}^N))^T \hat{\nu} \;=\; 0
\end{aligned}$$

where

$$\hat{\lambda}^j = \frac{\bar{\lambda}^j}{w_j}, \qquad \hat{\mu}^j = \frac{\bar{\mu}^j}{w_j}, \qquad \hat{\nu} = \bar{\nu} \tag{11}$$

These conditions are exactly the discretization of the continuous-time transversality conditions.

Also, due to the relaxation on the first and the last nodes,

$$\begin{aligned}
\frac{\partial L}{\partial \bar{x}^0} &= -\sum_{j=0}^{N} D_{j0}\bar{\lambda}^j + f_x^T(\bar{x}^0, \bar{u}^0)\bar{\lambda}^0 + F_x(\bar{x}^0, \bar{u}^0)w_0 + h_x^T(\bar{x}^0, \bar{u}^0)\bar{\mu}^0 - \bar{\lambda}^0/w_0 \;=\; 0 \\
\frac{\partial L}{\partial \bar{x}^N} &= -\sum_{j=0}^{N} D_{jN}\bar{\lambda}^j + f_x^T(\bar{x}^N, \bar{u}^N)\bar{\lambda}^N + F_x(\bar{x}^N, \bar{u}^N)w_N + h_x^T(\bar{x}^N, \bar{u}^N)\bar{\mu}^N + \bar{\lambda}^N/w_N \;=\; 0
\end{aligned}$$

The remaining part of KKT conditions remain the same as in Problem $B^{N\lambda}$.

It is important to note that this approach is not a new PS method but an alternative implementation of the same PS method.

Now, introducing a new variable

$$\hat{D} \;=\; -W^{-1}D^{T}W + W^{-1}\Delta \tag{12}$$

where W is a diagonal matrix defined as

$$W \;=\; diag(w_0, w_1, \ldots, w_N)$$

and

$$\Delta \;=\; \begin{bmatrix} -1 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ & & \vdots & \\ 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 1 \end{bmatrix}$$

Under notations $\hat{D}$ and (11), the KKT conditions of the modified Problem B^N are summarized in the following.

$$(\mathsf{KKT}) \quad \begin{cases} \sum_{j=0}^{N} \hat{D}_{kj}\hat{\lambda}^{j} + f_x^T(\bar{x}^k, \bar{u}^k)\hat{\lambda}^k + F_x(\bar{x}^k, \bar{u}^k) + h_x^T(\bar{x}^k, \bar{u}^k)\hat{\mu}^k \;=\; 0 \\ F_u(\bar{x}^k, \bar{u}^k) + f_u^T(\bar{x}^k, \bar{u}^k)\hat{\lambda}^k + h_u^T(\bar{x}^k, \bar{u}^k)\hat{\mu}^k \;=\; 0 \\ \hat{\mu}^k \cdot h(\bar{x}^k, \bar{u}^k) \;=\; 0, \qquad \bar{\mu}^k \;\geq\; 0 \\ \hat{\lambda}^0 + \dfrac{\partial E}{\partial x^0}(\bar{x}^0, \bar{x}^N) + (\dfrac{\partial e}{\partial x^0}(\bar{x}^0, \bar{x}^N))^T\hat{\nu} \;=\; 0 \\ -\hat{\lambda}^N + \dfrac{\partial E}{\partial x^N}(\bar{x}^0, \bar{x}^N) + (\dfrac{\partial e}{\partial x^N}(\bar{x}^0, \bar{x}^N))^T\hat{\nu} \;=\; 0 \end{cases}$$

Costate Computation

Comparing the KKT condition of the modified Problem B^N with the discretization of the continuous necessary conditions Problem $B^{\lambda N}$, we can easily identify a key difference: adjoint equation is discretized by $\hat{D}$ not the original differentiation matrix D. In the case of Legendre-Gauss-Lobatto PS method, the corresponding differentiation matrix satisfies the following condition[7]

$$\begin{aligned} w_i D_{ik} \;&=\; -w_k D_{ki}, \quad \text{if } k \;\neq\; i \\ D_{ii} \;&=\; 0, \qquad\quad \text{if } i \neq 1, N \end{aligned} \tag{13}$$

From (12) and (13), it is easy to see $\hat{D} = D$. Therefore, if LGL nodes are chosen, the adjoint equations are discretized by the same differentiation matrix.

When other types of nodes are chosen, $\hat{D}$ is, in general, not equal to D. In fact, the difference between $\hat{D}$ and D can be quite large. This prompts a need to study the property of the matrix $\hat{D}$. To this end, we let $\lambda^N(t)$ be a polynomial approximation of the continuous costate $\lambda(t)$, i.e.,

$$\lambda_i(t) \;\approx\; \lambda_i^N(t) = \sum_{k=0}^{N} \bar{\lambda}_i^k \phi_k(t), \qquad i = 1, \cdots, N_x$$

We have

$$\begin{aligned}
\hat{D}\cdot\bar{\lambda}_i^T &= W^{-1}\left(-D^T\cdot W\cdot\bar{\lambda}_i+\Delta\cdot\bar{\lambda}_i\right)\\
&= W^{-1}\left(-\begin{bmatrix}\dot{\phi}_0(t_0)w_0 & \dot{\phi}_0(t_1)w_1 & \cdots & \dot{\phi}_0(t_N)w_N\\ \dot{\phi}_1(t_0)w_0 & \dot{\phi}_1(t_1)w_1 & \cdots & \dot{\phi}_1(t_N)w_N\\ \vdots & \vdots & \cdots & \vdots\\ \dot{\phi}_N(t_0)w_0 & \dot{\phi}_N(t_1)w_1 & \cdots & \dot{\phi}_N(t_N)w_N\end{bmatrix}\bar{\lambda}_i+\begin{bmatrix}-\bar{\lambda}_i^0\\0\\\vdots\\0\\\bar{\lambda}_i^N\end{bmatrix}\right)\\
&= W^{-1}\begin{bmatrix}-\sum_{k=0}^N\dot{\phi}_0(t_k)\bar{\lambda}_i^k w_k-\bar{\lambda}_i^0\\-\sum_{k=0}^N\dot{\phi}_1(t_k)\bar{\lambda}_i^k w_k\\\vdots\\-\sum_{k=0}^N\dot{\phi}_{N-1}(t_k)\bar{\lambda}_i^k w_k\\-\sum_{k=0}^N\dot{\phi}_N(t_k)\bar{\lambda}_i^k w_k+\bar{\lambda}_i^N\end{bmatrix}
\end{aligned}$$

Since the weights w_k are used to approximate the integration in the discretization, we have

$$\sum_{k=0}^N\dot{\phi}_j(t_k)\bar{\lambda}_i^k w_k \approx \int_{-1}^1\dot{\phi}_j(t)\lambda_i^N(t)dt, \qquad \text{for all } j=0,1,\ldots,N$$

By integration by part formula,

$$\begin{aligned}
\hat{D}\cdot\bar{\lambda}_i^T &\approx W^{-1}\begin{bmatrix}-\int_{-1}^1\dot{\phi}_0(t)\lambda_i^N(t)dt-\lambda_i^N(t_0)\\-\int_{-1}^1\dot{\phi}_1(t)\lambda_i^N(t)dt\\\vdots\\-\int_{-1}^1\dot{\phi}_{N-1}(t)\lambda_i^N(t)dt\\-\int_{-1}^1\dot{\phi}_N(t)\lambda_i^N(t)dt+\lambda_i^N(t_N)\end{bmatrix}\\
&= W^{-1}\begin{bmatrix}\int_{-1}^1\phi_0(t)\dot{\lambda}_i^N(t)dt-\phi_0(1)\lambda_i^N(1)+\phi_0(-1)\lambda_i^N(-1)-\lambda_i^N(-1)\\\int_{-1}^1\phi_1(t)\dot{\lambda}_i^N(t)dt-\phi_1(1)\lambda_i^N(1)+\phi_1(-1)\lambda_i^N(-1)\\\vdots\\\int_{-1}^1\phi_{N-1}(t)\dot{\lambda}_i^N(t)dt-\phi_{N-1}(1)\lambda_i^N(1)+\phi_{N-1}(-1)\lambda_i^N(-1)\\\int_{-1}^1\phi_N(t)\dot{\lambda}_i^N(t)dt-\phi_N(1)\lambda_i^N(1)+\phi_N(-1)\lambda_i^N(-1)+\lambda_i^N(1)\end{bmatrix}
\end{aligned}$$

Since $\phi_j(t)$ are interpolating polynomials that satisfies $\phi_j(t_k)=0$, if $j\neq k$; and $\phi_j(t_k)=1$, if $j=k$, we have

$$\begin{aligned}
\hat{D}\cdot\bar{\lambda}_i^T &\approx W^{-1}\begin{bmatrix}\int_{-1}^1\phi_0(t)\dot{\lambda}_i^N(t)dt\\\vdots\\\int_{-1}^1\phi_N(t)\dot{\lambda}_i^N(t)dt\end{bmatrix}\approx W^{-1}\begin{bmatrix}\sum_{i=0}^N\phi_0(t_i)\dot{\lambda}_i^N(t_i)w_i\\\vdots\\\sum_{i=0}^N\phi_N(t_i)\dot{\lambda}_i^N(t_i)w_i\end{bmatrix}\\
&= W^{-1}\begin{bmatrix}\dot{\lambda}_i^N(t_0)w_0\\\vdots\\\dot{\lambda}_i^N(t_N)w_N\end{bmatrix}=\begin{bmatrix}\dot{\lambda}_i^N(t_0)\\\vdots\\\dot{\lambda}_i^N(t_N)\end{bmatrix}\approx\begin{bmatrix}\dot{\lambda}_i(t_0)\\\vdots\\\dot{\lambda}_i(t_N)\end{bmatrix}
\end{aligned}$$

This shows that $\hat{D}$ does provide an approximation to the derivatives. It also shows that the accuracy of the approximation highly depends on the accuracy of the integration weights. If the grids are chosen so that the resulting weights can provide an accurate numerical integration, then the continuous adjoint equation can also be accurately and automatically discretized by the KKT conditions by way of a different differentiation matrix.

Remark 7 *To minimize the error in costate computation, we should choose grids points so that the resulting numerical integration scheme is as accurate as possible. Once the boundary points are fixed at -1 and + 1, the optimal grid is given by Legendre-Gauss-Lobatto points and not Gauss-Radau or Gauss points. In the case of Chebyshev-Gauss-Lobatto grids, the corresponding Clenshaw-Curtis weights provides accurate integration. Indeed, Trefethen[22] provides an excellent analysis on the accuracy of Clenshaw-Curtis integration; in particular, he notes that Clenshaw-Curtis integration is practically as good as Gauss quadrature integration. Therefore, CGL grids are also a good choice for the purpose of both primal and dual state computation.*

NUMERICAL EXAMPLES

Consider the following example from Ref.[14].

Minimize $J[x(\cdot), u(\cdot)] = x(2)$, subject to

$$\begin{aligned} \dot{x}(t) &= u(t), \qquad t \in [0, 2] \\ x(0) &= 0, \qquad u(t) \geq -1 \end{aligned}$$

The necessary conditions

$$\begin{aligned} \dot{\lambda}^*(t) &= 0, \qquad \lambda^*(2) = 1 \\ \lambda^*(t) - \mu^*(t) &= 0 \\ \mu^*(t)(-u^*(t) - 1) &= 0, \qquad \mu^*(t) \geq 0 \end{aligned}$$

uniquely determine the optimal solution as

$$\begin{aligned} x^*(t) &= -t, \qquad u^*(t) = -1, \\ \lambda^*(t) &= 1, \qquad \mu^*(t) = 1. \end{aligned}$$

To test the idea on arbitrary grids, we choose uniformly distributed nodes with $N = 10$. With this nodes distribution, it is easy to compute the corresponding differentiation matrix and weights using the aforementioned formula. We solve the problem using the presented PS method. The result is showing in Fig. 2. Both primal and dual variables appear to converge.

Given the previous converged numerical results, it is natural to assume that when the size of the grid increases we will get a better solution. In fact, for a uniform grid, this is not true. Fig.3 is a simulation result with N=12. Clearly, none of the discrete solutions are correct. This is exactly due to the violation of Assumption 2. For uniform grids, it is known that when $N \geq 11$, at least one of the weight is negative and the discrete approximation of the integration diverges. This is why uniform grids with high order polynomial approximation is not suggested for numerical computation. However, it does not mean high order polynomial should be avoided. The remedy is well known: use quadrature types of nodes with dense distribution around two end points. Such grids not only

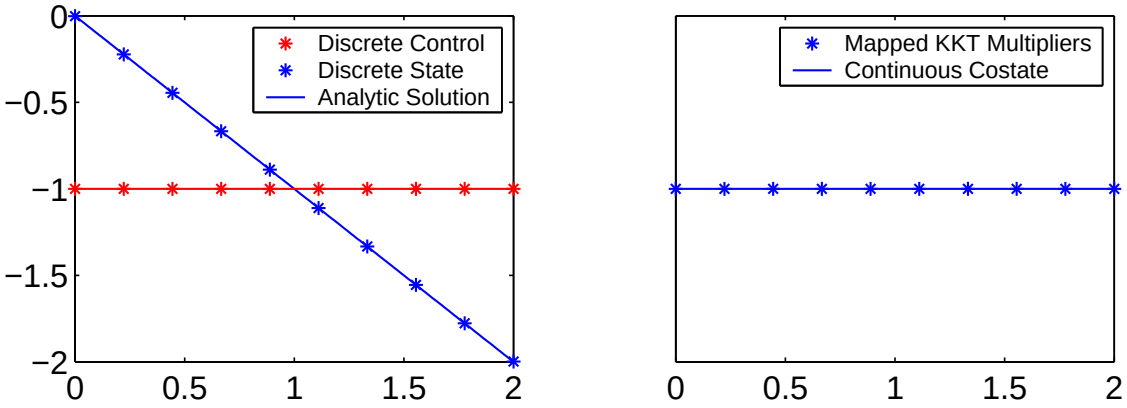

Figure 2. Discrete solution on uniform grids (N=10) with primal-only closure condition.

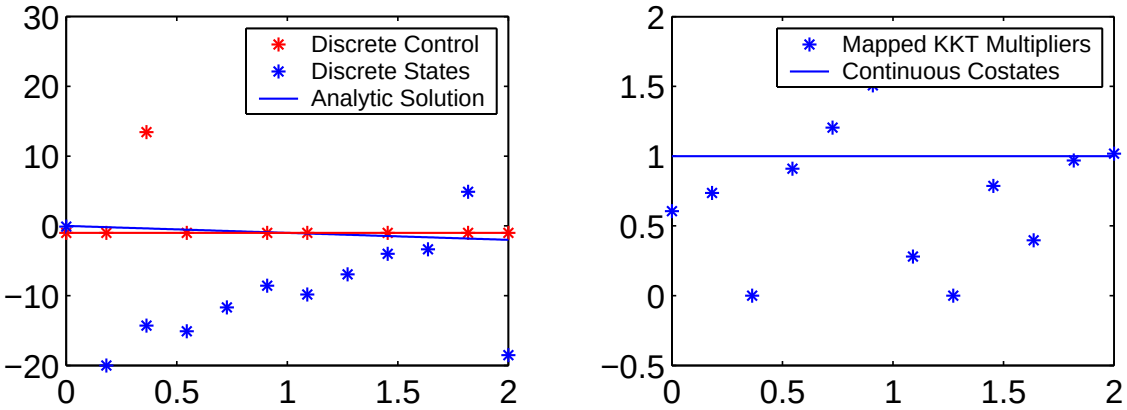

Figure 3. Discrete solution on uniform grids (N=12) with primal-only closure condition.

overcome the problem associated to uniform grids, but also greatly improve the accuracy. In the following, we demonstrate the proposed PS scheme on Chebyshev grids.

Consider a minimum fuel orbit transfer problem.

$$\begin{cases} \text{Minimize} & J[\cdot] = \int_0^{20} u_r^2 + u_t^2 dt \\ \text{Subject to} & \dot{r} = v_r \\ & \dot{\theta} = \frac{v_t}{r} \\ & \dot{v}_r = \frac{v_t^2}{r} - \frac{1}{r^2} + u_r \\ & \dot{v}_t = -\frac{v_r v_t}{r} + u_t \\ & |u_r| \leq 0.05; \;\; |u_t| \leq 0.05 \\ & (r(0), v_r(0), v_t(0)) = (1, 0, 1) \\ & (r(t_f), v_r(t_f), v_t(t_f)) = (4, 0, 0.5) \end{cases}$$

where r is the radial distance, θ is the true anomaly, v_r and v_t are the velocities on radial and transverse directions, u_r and u_t are the radial and transverse thrust. The equations are presented in normalized units.

Fig.4 shows the optimal trajectory and optimal control computed by Chebyshev PS method with 64 nodes. Also shown in Fig. 4 are the states obtained from a numerical (RK4/5) propagation of the

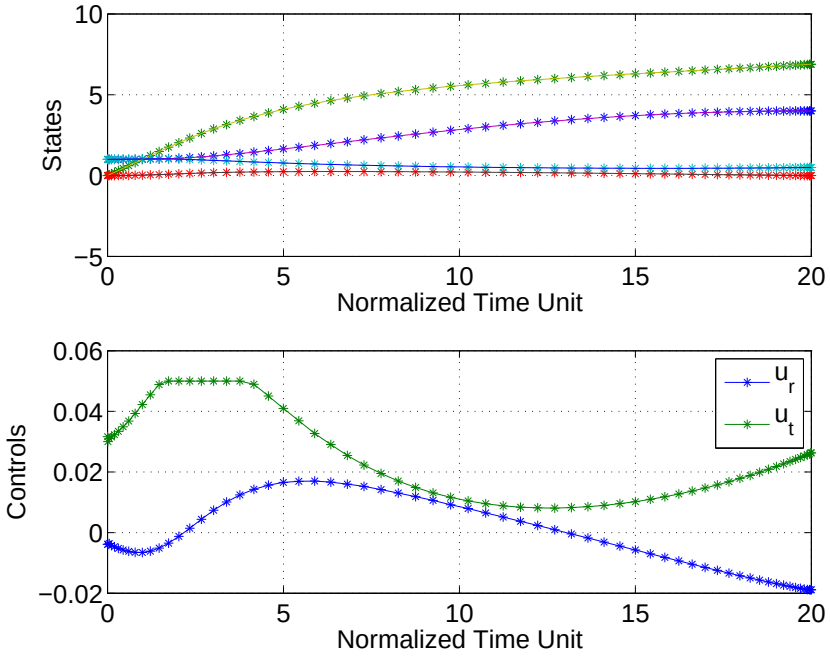

Figure 4 Optimal primal solutions. The solid lines are the propagated trajectories generated by linear interpolated controls.

discrete-time optimal controller. Clearly, the discrete optimal states match the propagated trajectory very accurately, which numerically demonstrates the feasibility and accuracy of the discrete optimal solution.

To further demonstrate the optimality, we plot out the costates in Fig.5. These costates are computed using the proposed primal-Only closure conditions. Based on the Minimum Principle, it is straightforward to derive the following adjoint equations

$$\begin{cases} \dot{\lambda}_r &= \lambda_\theta \frac{v_t}{r^2} + \lambda_{vr} \frac{vt^2}{r^2} - \frac{2\lambda_{vr}}{r^3} - \lambda_{vt} \frac{v_r v_t}{r^2} \\ \dot{\lambda}_\theta &= 0 \\ \dot{\lambda}_{vr} &= -\lambda_r + \frac{\lambda_{vt} v_t}{r} \\ \dot{\lambda}_{vt} &= -\frac{\lambda_\theta}{r} - \frac{2\lambda_{vr} v_t}{r} + \frac{\lambda_{vt} v_r}{r} \end{cases}$$

The fact that λ_θ is a constant is clearly shown in Fig.5. Also, by Minimum Principle, Hamiltonian should be constant 0. The discrete Hamiltonian shown in Fig.5 agrees with this fact. Thus, the optimality of the computed solution is verified through costates mapping.

This example is also solved using Legendre-Gauss-Lobatto grid with the same number of nodes (N=64). The PS scheme with primal-only closure condition yields the same solution as Chebyshev case.

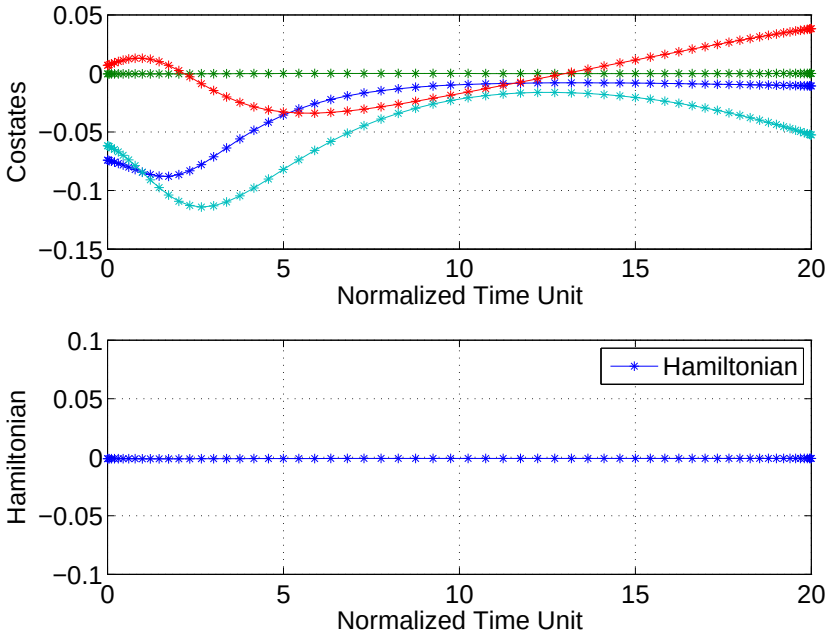

Figure 5. Costates by Chebyshev PS methods.

REFERENCES

[1] Kang, W. and Bedrossian, N., "Pseudospectral Optimal Control Theory Makes Debut Flight – Saves NASA $1M in Under 3 hrs," *SIAM News*, Vol. 40, No. 7, September 2007, Page 1.

[2] G. Elnagar, M. A. Kazemi and M. Razzaghi, The pseudospectral Legendre method for discretizing optimal control problems, *IEEE Trans. Automat. Contr.* Vol. 40, pp. 1793-1796, 1995.

[3] S. W. Paris, J. P. Riehl, and W. K. Sjaw, "Enhanced Procedures for Direct Trajectory Optimization Using Nonlinear Programming and Implicit Integration," AIAA/AAS Astodynamics Specialist Conference and Exhibit, 21-24 August 2006, Keystone, CO, AIAA 2006-6309.

[4] P. Lu, H. Sun and B. Tsai, Closed-loop endoatmospheric ascent guidance, *Journal of Guidance, Control and Dynamics*, Vol. 26, No. 2, pp.283-294, 2003.

[5] S. I. Infeld, and W. Murray, Optimization of stationkeeping for a libration point mission, *AAS Spaceflight Mechanics Meeting*, Maui, HI, AAS 04-150, February 2004.

[6] Q. Gong, W. Kang and I. M. Ross, A pseudospectral method for the optimal control of constrained feedback linearizable systems, *IEEE Trans. on Automatic Control*, Vol. 51, No. 7, pp. 1115-1129, 2006.

[7] F. Fahroo and I. M. Ross, Costate estimation by a Legendre pseudospectral method, *AIAA Journal of Guidance, Control and Dynamics*, Vol. 24, No. 2, pp. 270-277, 2001.

[8] P. Williams, B. Lansdorp and W. Ockels, Optimal crosswind towing and power generation with tethered kites, *AIAA Journal of Guidance, Control, and Dynamics*, Vol. 31 no. 1, pp. 81-93, 2008.

[9] R. E. Stevens and W. Wiesel, Large time scale optimal control of an electrodynamic tether satellite, *Journal of Guidance, Control and Dynamics*, Vol. 32, No. 6, pp.1716-1727, 2008.

[10] Bedrossian, N., Bhatt, S., Lammers, M., Nguyen, L., and Zhang, Y., First ever flight demonstration of zero propellant maneuver attitude control concept, *AIAA Guidance, Navigation and Control Conference*, 2007. AIAA-2007-6734.

[11] F. Fahroo and I. M. Ross , Advances in pseusospectral methods for optimal control, *AIAA Guidance, Navigation, and Control Conference*, AIAA Paper 2008-7309, Honolulu, Hawaii, August 2008.

[12] F. Fahroo and I. M. Ross, Pseudospectral methods for infinite-horizon nonlinear optimal control problems, *Journal of Guidance, Control and Dynamics*, Vol. 31, No. 4, pp.927-936, 2008.

[13] Fahroo, F. and Ross, I. M., On Discrete-Time Optimality Conditions for Pseudospectral Methods, *AIAA/AAS Astrodynamics Conference*, Keystone, CO, August 2006. AIAA-2006-6304

[14] Q. Gong, M. Ross, W. Kang, F. Fahroo, Connections between the covector mapping theorem and convergence of pseudospectral methods for optimal control, *Computational Optimization and Applications*, Vol. 41, pp. 307-335, 2008.

[15] W. Kang, Q. Gong, and I. M. Ross, On the convergence of nonlinear optimal control using pseudospectral methods for feedback linearizable systems, *International Journal of Robust and Nonlinear Control*, Vol. 17, Issue 14, pp. 1251-1277, 2007.

[16] W. Kang, I. M. Ross, Q. Gong, Pseudospectral optimal control and its convergence theorems, Analysis and Design of Nonlinear Control Systems - In Honor of Alberto Isidori, Alessandro Astolfi and Lorenzo Marconi eds., Springer, pp. 109-124, 2008.

[17] Williams, P., Jacobi Pseudospectral Method for Solving Optimal Control Problems, *Journal of Guidance, Control and Dynamics*, Vol. 27, No. 2, pp.293-297, 2004.

[18] B. S. Mordukhovich, *Variational Analysis and Generalized Differentiation, I: Basic Theory*, vol. 330 of Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences] Series, Springer, Berlin, 2005.

[19] B. S. Mordukhovich, *Variational Analysis and Generalized Differentiation, II: Applications*, vol. 331 of Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences] Series, Springer, Berlin, 2005.

[20] C. Canuto, M. Y. Hussaini, A. Quarteroni, and T. A. Zang, *Spectral Methods in Fluid Dynamics,* Springer Verlag, New York, 1988.

[21] E. Polak, *Optimization: Algorithms and Consistent Approximations*, Springer-Verlag, Heidelberg, 1997.

[22] L. N. Trefethen, Is Gauss quadrature better than Clenshaw-Curtis? *SIAM Review*, Vol. 50, No. 1, pp 67-87, 2008.